Time Series Data Analysis Using R

- Introduction to R
- Getting Started - Using RStudio IDE
  - R 3.2.x
  - RStudio 0.99.xxx
- On Line Data Resources
  - quantmod
  - Quandl

Time Series Data

- Economic Time Series Data
  - GDP, CPI, Oil Price, Climate Change
  - Interest Rates, Exchange Rates
- Financial Time Series Data
  - S&P 500, VIX (Fear Index)
  - International Stock Markets
- High Frequency Time Series

$$\{y_t\} = \{\ldots, y_1, y_2, \ldots, y_T, \ldots\}$$

Time Series Data

- Data Exploration
  - Using Graphs
    - Trend $\{y_t\} = \{\ldots, y_1, y_2, \ldots, y_T, \ldots\}$
    - Seasonality
  - Transformation: Lag, Difference
  - Time Series Decomposition
    - Additive Components $y_t = m_t + s_t + \varepsilon_t$
    - Multiplicative Components $\log(y_t) = m_t + s_t + \varepsilon_t$

Time Series Data

- Data Analysis
  - Hypothesis Testing
    - Durbin-Watson
    - Box-Pierce / Ljung-Box
    - ACF/PACF
  - Exponential Smoothing
  - ARIMA Model
    - Estimation
    - Forecasting

Exponential Smoothing

$$y_t = m_t + s_t + \varepsilon_t \ or \ \log(y_t) = m_t + s_t + \varepsilon_t$$

$$trend: m_t = a_t + b_t \ (level + slope)$$

$$seasonal: s_t \quad random: \varepsilon_t$$

– Simple Exponential Smoothing (EWMA)

$$b_t = 0, s_t = 0$$

– Holt Exponential Smoothing

$$s_t = 0$$

– Holt-Winters Exponential Smoothing

Exponential Smoothing

- Simple Exponential Smoothing (EWMA)

$$y_{t+1} = a_t + b_t + s_t + \varepsilon_t$$

$$b_t = 0, s_t = 0$$

  – Smoothing Equation

$$a_t = \alpha y_t + (1-\alpha) a_{t-1}, \quad 0 < \alpha < 1$$

$$= \alpha y_t + \alpha(1-\alpha) y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \ldots$$

  – Forecast Equation

$$\hat{y}_{t+1|t} = a_t, \quad t = 1, 2, \ldots, n$$

$$\hat{y}_{n+h|n} = a_n, \quad h = 1, 2, \ldots$$

Exponential Smoothing

- Simple Exponential Smoothing (EWMA)
  - State Space Representation

$$y_t = a_{t-1} + e_t \quad (observation\ equation)$$

$$a_t = a_{t-1} + \alpha e_t \quad (state\ equation)$$

$$a_1 = y_1 \quad initialization$$

$$a_2 = a_1 + \alpha(y_2 - a_1)$$

$$\ldots$$

$$a_n = a_{n-1} + \alpha(y_n - a_{n-1})$$

$$\hat{\alpha} = \min_{\alpha} \arg \sum_{t=1}^{n} e_t^2$$

Exponential Smoothing

- Holt Exponential Smoothing

$$y_t = a_{t-1} + b_{t-1} + e_t$$

$$a_t = a_{t-1} + b_{t-1} + \alpha e_t$$

$$b_t = b_{t-1} + \alpha\beta e_t$$

$$(\hat{\alpha}, \hat{\beta}) = \min_{\alpha,\beta} \arg \sum_{t=1}^{n} e_t^2$$

$$where\ e_t = y_t - (a_{t-1} + b_{t-1})$$

Exponential Smoothing

- Holt Exponential Smoothing
  - Smoothing Equation

$$a_t = \alpha y_t + (1-\alpha)(a_{t-1} + b_{t-1}), \quad 0 < \alpha < 1$$

$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}, \quad 0 < \beta < 1$$

$$s_t = 0$$

  - Forecast Equation

$$\hat{y}_{t+1|t} = a_t + b_t, \quad t = 1, 2, \ldots, n$$

$$\hat{y}_{n+h|n} = a_n + hb_n, \quad h = 1, 2, \ldots$$

Exponential Smoothing

- Holt-Winters Exponential Smoothing

$$y_t = a_{t-1} + b_{t-1} + s_{t-p} + e_t$$

$$a_t = a_{t-1} + b_{t-1} + \alpha e_t$$

$$b_t = b_{t-1} + \alpha \beta e_t$$

$$s_t = s_{t-p} + (1-\alpha)\gamma e_t$$

$$(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = \min_{\alpha,\beta,\gamma} \arg \sum_{t=1}^{n} e_t^2$$

$$where\; e_t = y_t - (a_{t-1} + b_{t-1} + s_{t-p})$$

Exponential Smoothing

- Holt-Winters Exponential Smoothing
  - Smoothing Equation

$$a_t = \alpha(y_t - y_{t-p}) + (1-\alpha)(a_{t-1} + b_{t-1}), \quad 0 < \alpha < 1$$

$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}, \quad 0 < \beta < 1$$

$$s_t = \gamma(y_t - a_{t-1}) + (1-\gamma)s_{t-p}, \quad 0 < \gamma < 1$$

  - Forecast Equation

$$\hat{y}_{t+1|t} = a_t + b_t + s_{t+1-p}, \quad t = 1, 2, \ldots, n$$

$$\hat{y}_{n+h|n} = a_n + hb_n + s_{n+h-p}, \quad h = 1, 2, \ldots \leq p$$

Time Series Forecasting

- h-Step Ahead Forecast at t = n

$$\hat{y}_{n+h|n} = a_n + hb_n + s_{n+h-p}, \quad h = 1, 2, \ldots \leq p$$

- Forecast Error (if y_{n+h} is known)

$$\hat{\varepsilon}_{n+h|n} = y_{n+h} - \hat{y}_{n+h|n}$$

Time Series Forecasting

- One-Step Ahead Forecast at t = p, p+1,…

$$\hat{y}_{t+1|t} = a_t + b_t + s_{t+1-p}$$

$$\hat{y}_{1|0} = a_0 + b_0 + s_{1-p}$$

$$\hat{y}_{2|1} = a_1 + b_1 + s_{2-p}$$

$$\ldots$$

$$\hat{y}_{p|p-1} = a_{p-1} + b_{p-1} + s_0$$

$$\hat{\varepsilon}_{t+1|t} = y_{t+1} - \hat{y}_{t+1|t}$$

$$MAE = mean(|\,\hat{\varepsilon}_{t+1|t}\,|)$$

$$MAPE = 100\ mean(|\,\hat{\varepsilon}_{t+1|t} / y_{t+1}\,|)$$

$$MSE = mean(\hat{\varepsilon}^2_{t+1|t})$$

$$RMSE = \sqrt{mean(\hat{\varepsilon}^2_{t+1|t})}$$

$$RMSPE = 100\sqrt{mean[(\hat{\varepsilon}_{t+1|t} / y_{t+1})^2]}$$

Example 1

- China Shanghai Common Stock
  - High Frequency Daily Index
  - Monthly Index Time Series
    - Trend, Seasonality
  - Monthly Log-Return and Volatility
  - Time Series Decomposition
  - Exponential Smoothing and Forecasting

Example 2

- GDP and GDP Growth
  - GDP Quarterly Time Series
    - Trend, Seasonality
  - GDP Growth
    - Simple vs Compound Growth Rate
    - Quarterly Growth vs Annual Growth
  - Time Series Decomposition
  - Smoothing and Forecasting